

# Model-based ML for Retrospective Event Detection

Joan Capdevila<sup>\*†</sup>, Jesús Cerquides<sup>‡</sup>, and Jordi Torres<sup>\*†</sup>

<sup>\*</sup> Barcelona Supercomputing Center (BSC), Barcelona, Spain

<sup>†</sup> Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

<sup>‡</sup> IIIA-CSIC, Bellaterra, Spain

E-mail: {joan.capdevila, jordi.torres}@bsc.es, cerquide@iia.csic.es

**Keywords**—*Model-based Machine Learning, Probabilistic Models, Variational Inference, Social Computing, Twitter.*

## I. INTRODUCTION

The problem of event detection in Twitter has lately attracted the interest of several distinct communities ranging from data miners, spatial statisticians to machine learners. Each community has proposed tailored solutions to a problem which has multiple definitions in the literature. This has led to a myriad of techniques which address similar but slightly different problems [1]. In this paper, we focus on Retrospective Event Detection (RED) from geo-located tweets. RED seeks to identify and characterize groups of tweets which belong to a past unseen event [2]. This problem can be of interest to news agencies, city councils or geo-marketing companies.

One common approach to RED formulates the problem as one of clustering with noise. That is to say that event-related tweets are associated to event clusters while non-event tweets are assigned to a noise or background component. For example, Capdevila et al. presented Tweet-SCAN in [3], a technique for RED that extends the popular clustering technique called DBSCAN [4] to cope with geo-located tweets. However, this type of approach to machine learning constraints the solution to the assumptions of the chosen technique. For example, Tweet-SCAN suffers from the same weakness than DBSCAN which is that all event-related clusters must have the same minimum tweet density. Addressing this shortcoming within this approach would certainly require to replace DBSCAN with OPTICS or another technique. In most cases, this type of changes prevent us to reuse parts of the original solution when trying to incorporate more complex assumptions.

Model-based Machine Learning [5] offers an alternative approach which separates assumptions from computation and tasks. It enables to explicitly specify the assumptions in a compact modeling language, then define an inference algorithm to learn the model from data and finally carry out the tasks as predictions on the trained model. Moreover, Blei [6], inspired by the work of George E.P. Box, proposed to close the loop with criticism, so that assumptions can be revised when the solution does not meet the requirements. In this paper, we present WARBLE [7], a model-based solution that specifies the assumptions through a probabilistic graphical model, defines a variational inference algorithm to learn the model from “La Mercè” data set and performs event detection through a Maximum a Posteriori (MAP) query on the event assignment variables. We concluded in criticism section presenting some results and conclusions of this approach.

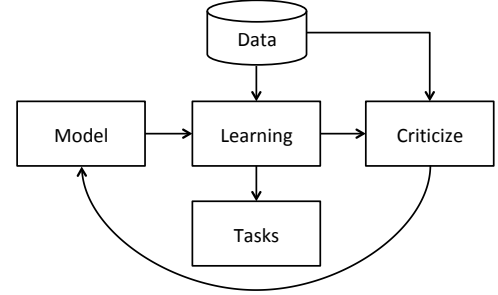


Fig. 1: Box’s loop applied to Model-based Machine Learning [6].

## II. THE WARBLE SOLUTION

### A. The Probabilistic Model

The WARBLE model is a heterogeneous mixture model with two types of mixture components. On the one hand, the event components represent the different events in terms of their spatio-temporal and textual features. On the other hand, the non-event component depicts the non-event tweets in its steady conditions. The generation process for event and non-event tweets can be described through the following generation processes,

Event tweet	Non-event tweet
$l_n \sim \text{Normal}(\mu_k, \Delta_k)$	$l_n \sim \text{Hist}(L_B)$
$t_n \sim \text{Normal}(\tau_k, \lambda_k)$	$t_n \sim \text{Hist}(T_B)$
For $m = 1 \dots M_n$ :	For $m = 1 \dots M_n$ :
$z_{n,m} \sim \text{Discrete}(\theta_k)$	$z_{n,m} \sim \text{Discrete}(\theta_K)$
$w_{n,m} \sim \text{Discrete}(\phi_{z_{n,m}:})$	$w_{n,m} \sim \text{Discrete}(\phi_{z_{n,m}:})$

where  $l_n$  and  $t_n$  are the tweet location and time which we assume to be normally distributed for event tweets. Non-event tweets follow a Histogram distribution that models the varying spatio-temporal tweet density in steady conditions, addressing one of the shortcomings of Tweet-SCAN [3].  $w_{n,:}$  correspond to the  $M_n$  words in the tweet, which are distributed according to a mixture of topics  $\phi$  for both type of tweets. The proposed model also incorporates some prior distribution over the means, precisions and proportions of the above distributions (see [7] for more details).

### B. The Variational Learning Algorithm

The Bayesian approach of learning consists in computing the posterior distribution over all the model unknowns. How-

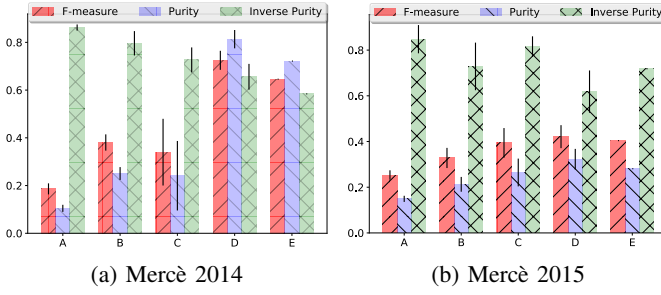


Fig. 2: (A) McInerney & Blei model (B) WARBLE w/o simultaneous topic-event learning (C) WARBLE w/o background model (D) Complete WARBLE (E) Tweet-SCAN

ever, this posterior distribution is intractable in the WARBLE model since it involves computing a normalizing constant which has to marginalize all latent variables and parameters. Approximation methods have been developed to bound this normalizing constant and solve a simpler optimization problem. Mean-field variational inference assumes a factorized distribution  $q(X|\eta)$  over all model parameters  $X = \{c, z, \pi, \tau, \lambda, \mu, \Delta, \theta, \phi\}$  and maximizes the evidence lower bound (ELBO) w.r.t the variational parameters  $\eta$

$$\text{ELBO}(\eta) = \mathbb{E}_{q(X|\eta)} [p(l, t, w, X; \Gamma)] - \mathbb{E}_{q(X|\eta)} [q(X|\eta)] \quad (1)$$

where  $\Gamma$  refers to the set of hyperparameters. Thanks to the mean-field approximation and the local conjugacy, we can build a coordinate descent algorithm with close-form updates for each variational parameter.

### C. The Event Detection Task

The event detection task can be seen as MAP query on the event assignment variables in which we assign the most probable mixture components (event or non-event) to each tweet. However, this inference task is also intractable and we need to resort to its variational approximation:

$$c^* = \underset{c}{\operatorname{argmax}} p(c|l, t, w; \Gamma) \approx \underset{c}{\operatorname{argmax}} q(c; \eta). \quad (2)$$

We note that this approach enables to formulate other tasks as queries to the probabilistic model. For example, we could ask for the most likely location for a tweet without geo-location.

### D. “La Mercè” data sets

We have crawled two datasets of tweets geo-located in the city of Barcelona during its local festivities in 2014 and 2015<sup>1</sup>. Several event-related tweets were manually tagged based on the festivities agenda. Tags were only used for evaluation purposes. Moreover, the histogram distributions for the spatio-temporal profiles for the non-event component were built from tweets generated during the days previous to the period of interest.

## III. CRITICISM

In Fig. 2, we compare the performance of WARBLE against other non-model-based techniques like Tweet-SCAN [3] and against other model-based approaches like McInerney & Blei

model [8]. We observe that WARBLE outperforms all techniques in both datasets in terms of F-measure. More importantly, we show how the iterative process of Model-based machine learning works. McInerney & Blei model did not distinguish between event and non-event tweets. By replacing their homogeneous mixture with a heterogeneous, we were able to improve accuracy from A to B. McInerney & Blei model did not perform simultaneous learning of topics-events. By doing so, we improve results from A to C. When combining both features into a more complex solution, we come up with WARBLE which has a performance as in D.

As we have seen, this approach to machine learning enables to be explicit about the assumptions, and hence criticize and improve different parts of the model. Moreover, it allows to decouple the model from inference and think of the computational aspects independently. This has pushed the field of *probabilistic programming* to develop software solutions which automatize inference for a given model.

## IV. ACKNOWLEDGMENT

This work is partially supported by Obra Social la Caixa. The extended abstract has been already published in [7].

## REFERENCES

- [1] F. Atefeh and W. Khreich, “A survey of techniques for event detection in twitter,” *Computational Intelligence*, vol. 31, no. 1, pp. 132–164, 2015.
- [2] Y. Yang, T. Pierce, and J. Carbonell, “A study of retrospective and on-line event detection,” in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1998, pp. 28–36.
- [3] J. Capdevila, J. Cerquides, J. Nin, and J. Torres, “Tweet-scan: An event discovery technique for geo-located tweets,” *Pattern Recognition Letters*, vol. 93, pp. 58 – 68, 2017, pattern Recognition Techniques in Data Mining.
- [4] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.
- [5] C. M. Bishop, “Model-based machine learning,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, p. 20120222, 2013.
- [6] D. M. Blei, “Build, compute, critique, repeat: Data analysis with latent variable models,” *Annual Review of Statistics and Its Application*, vol. 1, pp. 203–232, 2014.
- [7] J. Capdevila, J. Cerquides, and J. Torres, “Mining urban events from the tweet stream through a probabilistic mixture model,” *Data Mining and Knowledge Discovery*, Aug 2017.
- [8] J. McInerney and D. M. Blei, “Discovering newsworthy tweets with a geographical topic model,” *NewsKDD: Data Science for News Publishing workshop. Workshop in conjunction with KDD2014 the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2014.



**Joan Capdevila** is a Ph.D. student at the Computer Architecture Department at Universitat Politècnica de Catalunya (UPC) and associated to the Barcelona Supercomputing Center (BSC). He is supported by “la Caixa” Foundation through the UPC - “la Caixa” program. He has an Eng. degree in Telecommunications and a M.Sc. in Information and Communication Technologies by UPC. His research interest are in probabilistic modeling and approximate inference methods.

<sup>1</sup><https://github.com/jcapde/Twitter-DS/tree/master/MERCE/>